DOCUMENT RESUME

ED 052 294 VT 010 866

TITLE	Seven Studies on the Comparison of Live and Recorded Dictation Test Administration.
INSTITUTION	Manpower Administration (DOL), Washington, D.C. U.S. Training and Employment Service.
REPORT NO	USTES-RR-28
PUB DATE	Dec 69
NOTE	47p.
EDRS PRICE	EDRS Price MF-$0.65 HC-$3.29
DESCRIPTORS	Comparative Analysis, Comparative Testing, *Evaluation, *Performance Tests, *Stenography, *Testing

ABSTRACT

To compare the live method of administering dictation tests with the recorded method, the United States Training and Employment Service (USTES) conducted studies in cooperation with the State Employment Services of Alabama, Colorado, Minnesota, Mississippi, New York, and Utah. The procedures were similar in each state, with samples being broken into two or three groups and alternately given a live test and a recorded test. The bulk of the data supported the hypothesis that the recorded method produced essentially the same results as the live administration of the same tests. The correlations between the two methods were high enough to consider them equivalent in terms of ranking persons on their ability to take dictation. Some recommendations include: (1) Data on age, education, and experience should be collected for all subjects, (2) Scores on the new USTES Spelling Test should be available for all subjects, (3) Experimental groups should be similar on demographic variables and Spelling Test scores, and (4) Data should be available on equivalence of test forms used in research studies. (SB)

ED052294

December 1969 United States Training and Employment Service Test Research Report No. 28

Seven Studies on the Comparison of Live and Recorded Dictation Test Administration

VT010866

U.S. DEPARTMENT OF LABOR
MANPOWER ADMINISTRATION

U.S. DEPARTMENT OF HEALTH, EDUCATION & WELFARE
OFFICE OF EDUCATION
THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL OFFICE OF EDUCATION POSITION OR POLICY

ED052294

FOREWORD

Extensive research conducted under the Federal-State cooperative test research program in the Training and Employment Service has led to the development of many tools useful in vocational counseling and placement. These tools include aptitude tests, proficiency tests, and non-cognitive measures based on instruments such as interest inventories and biographical information blanks.

The purpose of this series of reports is to provide results of significant test research projects as they are completed. These reports will be of interest to users of the tests and to test research personnel in other organizations.

This report was prepared in the Division of Counseling and Testing Services of the United States Training and Employment Service by Eileen D. Haggerty and Marlin L. Ferral under the general direction of Anthony J. Fantaci, Chief of the Division, and Beatrice J. Dvorak, Assistant Chief of the Division.

U.S. Training and Employment Service

ACKNOWLEDGEMENT

The research described in this report was carried out by the U.S. Training and Employment Service in cooperation with the following six State Employment Services: Alabama, Colorado, Minnesota, Mississippi, New York, and Utah.

SEVEN STUDIES ON THE COMPARISON OF LIVE AND RECORDED DICTATION TEST ADMINISTRATION

Contents

	Page
Introduction	1
Study I (New York)	3
Study II (Colorado)	9
Study III (Colorado)	15
Study IV (Utah)	19
Study V (Alabama)	25
Study VI (Minnesota)	31
Study VII (Mississippi)	39
Summary of Findings for the Seven Studies	47
Recommendations	49
References	51

INTRODUCTION

The term "live" refers to the conventional method of administering dictation tests. The test administrator dictates standard material in person using a stopwatch. The term "recorded" means that the material to be dictated is on tape or phonograph recordings and is administered by using mechanical equipment connected to loudspeakers or earphones for sound production. The terms refer only to methods for administering dictated material and not to directions given examinees before dictation begins.

The recorded method should improve standardization of dictation test administration by eliminating expected variations in speed of reading, clarity of pronounciation, and voice quality which occur between different test administrators and within the same test administrator on different occasions. Five questions on method of dictation test administration were studied by State agencies affiliated with the U.S. Employment Service between 1959 and 1967. Each of the seven studies in this report concerned a number of the issues involved:

> (1) Whether there are differences in the mean scores obtained under live and recorded administration; (2) whether there are differences in the reliability of scores under the two methods of administration; (3) whether there are differences in the results obtained using different types of recorded administration--phonograph recordings or tape recordings, and loudspeaker or individual earphones; (4) whether recorded administration is regarded by examinees as less fair than live administration as a test of their skill in taking dictation; and (5) whether the same recording can be used with equal results in various parts of the country or different recordings are needed for various regions because of different regional speech patterns. (U.S. Department of Labor, Manpower Administration, March 1968)

The research was conducted using transcript dictation tests shown in the Guide to the Use of Typing, Dictation and Spelling Tests (U.S. Department of Labor, Bureau of Employment Security, December 1953). The Guide shows four-minute tests at 60 wpm and three-minute tests at 80 wpm. Raw scores can range from 0-240 errors, but the scores of the normative sample of 495 employed persons ranged from 0-88 errors. The distribution of error scores and corresponding letter grades for the normative sample is shown below

Errors	Percent of Sample	Letter Grade
0-1	7	A
2-5	24	B
6-14	38	C
15-52	24	D
53-88	7	E

These tests were used by the Employment Service prior to development of the true-false, separate answer sheet dictation tests now in use. The new USTES Dictation tests are different in text content and timing as well as method of response. It is assumed that research results obtained using transcription tests can be generalized to the use of the new true-false tests.

Empirical data on the equivalence of different transcription test forms used in these studies is not available. The forms were assumed to be equivalent by virtue of the similarity of their content and length.

The research designs, sample selection criteria, and statistical treatments were determined by the individual State agencies and vary accordingly. All State agencies used "Directions for Administration" shown in the Guide and followed the same general procedures for experimental test administration. Each test was preceded by a practice exercise at the same speed using the same type of administration (live or recorded). When recorded administration was used, only the dictation passages were recorded and directions were given live. The initial test and the second test were administered during one test administration period with a five-minute rest period between test administrations. Exceptions to standard procedures are noted in the separate studies.

STUDY I
(New York)

Summary

The first study was conducted by the New York employment security agency in 1959 to study mean differences between scores on the same form of the dictation test administered at 80 wpm by live and recorded methods to job applicants and to study the relationship of these scores. Two groups who had similar verbal ability test score means took Form No. 5 by both methods. The first group of 56 subjects took the test live first and recorded second. The second group of 50 subjects took the test recorded first and live second. Raw error scores were arranged in rank order and normalized scores were computed. Results of the study do not show any evidence that method of test administration affects the average level of dictation test score. Correlation coefficients between live and recorded administration ranged from .83 to .91.

Purpose

The general purpose of this study was to investigate the hypothesis that mode of test administration does not alter stenographic test performance. The specific purposes are listed, as follows:

(1) To investigate mean differences between dictation test scores on the same test form administered at 80 wpm by live and recorded methods to job applicants;

(2) To investigate the relationship of these live and recorded scores.

Procedure

Sample. The sample consisted of 106 applicants for stenographic jobs at the Commercial Office, 1 East 19th Street, New York City. Of these, 100 were females, and 6 were males. Testing started on January 20, 1959 and continued through February 3, 1959. The number tested on any one day ranged from 2 to 17. All applicants on these days who had not previously taken the specific form of the stenographic test being administered were included. Thus the sample included some applicants who might not have been ordinarily tested because of demonstrated recent experience as well as those usually tested about whom there might have been some question as to the degree of their stenographic ability. For each subject included in

the sample, scores were available on Verbal Ability Test CB1-J of B-1001 .

Test Administration. Form No. 5 was administered at 80 wpm to two groups. Fifty-six subjects in Group I took the test live first and recorded second; 50 subjects in Group II took the test recorded first and live second. The same examiner who gave the test in person recorded his administration on tape. The recordings were used with loudspeakers.

The 212 raw error scores obtained from testing 106 subjects twice were arranged in rank order and converted into normalized scores (Garrett, 1958). It was assumed that dictation test error scores are normally distributed.

Variables. One demographic variable, sex, was reported. Ordered variables treated in the analyses are listed, as follows:

For Group I (N=56)

1. Live normalized scores; first administration
2. Recorded normalized scores; second administration
3. Gain scores (from first to second administration)

For Group II (N=50)

4. Recorded normalized scores; first administration
5. Live normalized scores; second administration
6. Gain scores (from first to second administration)

For the total sample (N=106)

7. Verbal test scores
8. Live normalized scores
9. Recorded normalized scores

Statistical Treatment. Computations included means, standard deviations, and Pearson product-moment correlation coefficients. Means were compared by computing critical ratios for the standard error of the difference between means. Depending upon the data being compared, the formula selected was one for correlated means or one for uncorrelated means. The H test (Edwards, 1951) was applied to mean dictation test scores for the two experimental groups to determine whether the two groups sampled the same population on dictation test performance.

Results and Discussion

Results are shown in Tables I-1 and I-2. The correlations between dictation and verbal test scores are significant at the .01 level. Therefore, it would be appropriate to compare experimental groups matched on verbal ability. The means for verbal ability shown on Table I-1 were not significantly different at the .05 level.

Observations for dictation test data in Table I-1 are listed, as follows:

(1) Means for Group II are generally lower than means for Group I. Results of the H test show that dictation test population means for the two groups are not equal.

(2) Practice effect is suggested by the fact that mean performance is higher for the second test administration for each group.

(3) Means for total live administrations and total recorded administrations are not significantly different.

(4) Gains are very similar regardless of method of presentation used first.

Correlation coefficients between live and recorded scores shown in Table I-2 range from .83 to .91 indicating a very dependable relationship between scores obtained by two different methods of test administration.

Conclusion

The hypothesis that mode of administration does not alter stenographic test performance is supported.

TABLE I-1
(New York)

Means and Standard Deviations for Verbal Ability Scores, Normalized Rank Order Dictation Test Scores, and Dictation Gain Scores*

Tape recordings were made in New York and used with loudspeakers. The voice recorded was that of the examiner who gave the test by the live method.

Score	Group	N	M	SD
Verbal Test	I (Live first, recorded second)	56	26.1	7.5
Verbal Test	II (Recorded first, live second)	50	24.7	7.0
Dictation	I (Live scores)	56	50.1	20.2
Dictation	I (Recorded scores)	56	56.8	17.7
Dictation	II (Recorded scores)	50	43.0	17.3
Dictation	II (Live scores)	50	49.1	18.2
Dictation	Live scores--both groups	106	49.7	19.3
Dictation	Recorded scores--both groups	106	50.3	18.8
Dictation Gain	I (Live first, recorded second)	56	6.6	8.3
Dictation Gain	II (Recorded first, live second)	50	6.1	9.7

*Dictation test subjects took the same form (Form No. 5) at 80 wpm for first and second test administrations. Live raw scores ranged from 0-178 errors; recorded raw scores ranged from 0-209 errors.

TABLE I-2
(New York)

Correlation Coefficients between Verbal Scores (Form CB1-J of B-1001) and Dication Test Normalized Scores and between Live and Recorded Dictation Test Normalized Scores*

Tape recordings were made in New York and used with loudspeakers. The voice recorded was that of the examiner who gave the test by the live method.

Tests	Group	N	r**
Verbal and Dictation	All live administrations	106	.37
Verbal and Dictation	All recorded administrations	106	.36
Dictation	Group I (live first, recorded second)	56	.91
Dictation	Group II (recorded first, live second)	50	.85
Dictation	Total live vs. total recorded	106	.83

* Dictation test subjects took the same form (Form No. 5) at 80 wpm for first and second test administrations. Live raw scores ranged from 0-178 errors; recorded raw scores ranged from 0-209 errors.

**All r's shown are significantly different from zero at the .01 level.

STUDY II
(Colorado)

Summary

The second study was conducted by the Colorado employment security agency in 1960-61 to investigate the hypothesis that live and recorded methods of dictation test administration are equivalent. Subjects used individual earphones to hear phonograph recordings made by a Denver recording company with the trained voice of a local radio announcer. Dictation tests were given at 60 and 80 wpm in two sequences of presentation:

A. Live first, recorded second

B. Recorded first, live second

Different forms were used for first and second tests. There were 45 subjects in each experimental group. Analyses of mean error scores showed that order of presentation did not affect test performance. However, subjects made significantly more errors on the recorded method than on the live method. Correlations between live and recorded administrations ranged from .12 to .58. In this study the live and recorded methods are not equivalent. There are two reasons for assuming that there were problems with the sound production for the recorded method. There were 16 individuals, most of whom had scored well within the range of the accuracy norms (0-88 errors) on the live test, who missed all of the transcription on the recorded test (240 errors). Many subjects complained about using earphones. The results of this study, therefore, are not conclusive.

Purpose

The general purpose of this study was to investigate the hypothesis that live and recorded methods of test administration are equivalent. Specific purposes are listed, as follows:

1. To investigate mean differences between error scores on different dictation test forms administered (a) at 60 wpm and (b) at 80 wpm by live and recorded methods.

2. To investigate the relationship of these dictation test scores.

Procedure

Sample. Persons filing application for employment in stenographic positions were asked to volunteer for this study. A few business college students were also included. Data were incomplete for some of the 258 persons tested, and a few cases were eliminated at random to create four groups of equal size. The final sample is composed of 180 subjects with 45 subjects in each of the four groups.

Test Administration. Test forms used are listed, as follows:

At 60 wpm :

Live administration--Form No. 101
Recorded administration--Forms No. 102 and 103

At 80 wpm :

Live administration--Forms No. 5, 12, and 13
Recorded administration--Forms No. 10 and 11

The dictation tests were recorded by a Denver recording company using a local radio announcer who was also a teacher of diction. Phonograph recordings with individual earphones for sound production were used for testing. Subjects complained that recorded material was not dictated in phrases. Many subjects complained about devices in their ears.

Dictation for each speed was administered in two sequences:

A. live first, recorded second and

B. recorded first, live second

Variables. All ordered variables are error scores obtained at 60 and 80 wpm for four groups each having 45 subjects. Variables are listed, as follows:

At 60 wpm :

Sequence A
1. Live scores (first)
2. Recorded scores (second)

Sequence B
3. Recorded scores (first)
4. Live scores (second)

At 80 wpm :

Sequence A
5. Live scores (first)
6. Recorded scores (second)

Sequence B
7. Recorded scores (first)
8. Live scores (second)

Statistical Treatment. Means, standard deviations, and Pearson product-moment correlation coefficients were computed. Means were compared by computing critical ratios for the standard error of the difference between means. Depending upon the comparison, the formula selected was one for correlated means or one for uncorrelated means.

Results and Discussion

Error distributions are postively skewed for live and recorded tests. However, the range of scores is 0-181 for live scores and 0-240 for recorded scores. Sixteen individuals with scores of 240 errors (the maximum number possible) did not have any correct transcription on the recorded test. The accuracy of these measurements is questionable especially in view of the fact that most of these same individuals performed well within the normal 0-88 error range on the live test.

Means and standard deviations are shown in Table II-1. For live and recorded data combined, the A and B groups are not significantly different. However, the differences between live and recorded scores are significant. Subjects made more errors on recorded administration regardless of order of test administration. Standard deviations were larger for recorded test scores.

Table II-2 shows correlations ranging from .12 to .58 between live and recorded test scores. The correlation of .12 is not significantly different from zero.

Conclusions

The data do not support the hypothesis that live and recorded methods of test administration are equivalent. However, complaints of examinees about earphones, the occurence of 16 recorded scores at 240 errors, and the large standard deviations for recorded scores suggest the possibility of mechanical problems with sound production for the recorded method. The results of the study, therefore, are not conclusive.

TABLE II-1
(Colorado)

Means and Standard Deviations for Dictation Test Error Scores * Administered by Live and Recorded Methods at 60 and 80 wpm for Each Sequence of Presentation (N's=45)

Recordings were made by a Denver recording company with the trained voice of a local radio announcer. Phonograph recordings with individual earphones were used for test administration.

WPM	Sequence A				Sequence B			
	Live First		Recorded Second		Live Second		Recorded First	
	M	SD	M	SD	M	SD	M	SD
60	32.6	44.0	69.4	75.2	26.5	42.7	85.7	79.6
80	14.0	22.2	43.0	71.1	21.3	45.4	30.1	50.8

*Different test forms were used and are listed below with test score ranges:

At 60 wpm , live--Form No. 101 (0-181)
At 60 wpm , recorded--Forms No. 102 and 103 (0-240)
At 80 wpm , live--Forms No. 5, 12, and 13 (0-102)
At 80 wpm , recorded--Forms No. 10 and 11 (0-240)

TABLE II-2
(Colorado)

Correlation Coefficients for Dictation Test Error Scores* Administered by Live and Recorded Methods at 60 and 80 wpm for each Sequence of Presentation (N's=45)

Recordings were made by a Denver recording company with the trained voice of a local radio announcer. Phonograph recordings with individual earphones were used for test administration.

WPM	r (Sequence A)	r (Sequence B)
60	.58**	.12
80	.52**	.51 **

* Test forms and score ranges are listed as follows:

At 60 wpm , live--Form No. 101 (0-181)
At 60 wpm , recorded--Forms No. 102 and 103 (0-240)
At 80 wpm , live--Forms No. 5, 12, and 13 (0-102)
At 80 wpm , recorded--Forms No. 10 and 11 (0-240)

**Significant at the .01 level

14/-13-

STUDY III
(Colorado)

Summary

The Colorado employment security agency conducted Study III in 1966-67 to investigate the comparability of three methods of dictation test administration: live, phonograph recordings used with earphones, and phonograph recordings used with loudspeakers. Three groups of local office applicants were selected on the basis of identical dictation test scores on Form No. 5 of the dictation test administered live at 80 wpm. Each group had 26 subjects. Each group took Form No. 10 second by a different method of administration. Recordings were made in Colorado. In general, the data support the hypothesis that live and recorded administration methods are equivalent. However, the results are inconclusive because the procedure of selecting small groups of subjects on the basis of identical dictation test scores resulted in selection of three relatively high scoring groups (range 0-22). The second test scores were subject to regression toward the population mean which made them generally lower and more variable.

Purpose

The purpose of this study was to investigate the comparability of three methods of dictation test administration: live administration, phonograph recordings with individual earphones, and phonograph recordings with loudspeakers.

Procedure

Sample. From 235 local office applicants tested with the same test by the live method at 80 wpm, three groups of subjects were selected so that the dictation test score distribution would be perfectly equated. The final sample is composed of three groups each consisting of 26 subjects or a total number of 78 cases.

Test Administration. All subjects took Form No. 5 first and Form No. 10 second at 80 wpm. Form No. 10 was administered by three different methods, as follows:

Group 1--live
Group 2--phonograph recordings with earphones
Group 3--phonograph recordings with loudspeakers

Recordings were made in Colorado.

Variables. There were 26 cases for each of four ordered variables, as follows:

1. Error scores on first test (Form 5) administered live to all subjects in the study.

2. Error scores on second test (Form 10) administered live to Group 1.

3. Error scores on second test (Form 10) administered with earphones to Group 2.

4. Error scores on second test (Form 10) administered with loudspeakers to Group 3.

Statistical Treatment. Kolmogorov-Smirnov confidence bands for cumulative frequency were drawn for the population on Test No. 5 error scores (Walker and Lev, 1953). Means and standard deviations were computed for all variables. Means were compared by computing the t ratio for the standard error of the difference between correlated means (Edwards, 1951). Pearson product-moment correlation coefficients were computed between error scores on Tests No. 5 and 10.

Results and Discussion

All of the distributions fall within the .99 confidence band for the population represented by error scores on Test No. 5.

Means, standard deviations and ranges of error scores are shown in Table III-1. The difference between the mean of 7.1 for the initial live administration and 16.0 for the loudspeaker method is significant at the .05 level. None of the other mean differences is statistically significant for 26 cases.

Correlation coefficients between error scores on Test No. 5 and Test No. 10 are shown in Table III-2. Correlations range from .42 significant at the .05 level to .83 significant at the .01 level.

Conclusion

The findings on three methods are inconclusive. The procedure of selecting small groups of subjects on the basis of identical dictation test scores made second test scores subject to regression effects which tend to obscure true relationships and differences.

TABLE III-1
(Colorado)

Means, Standard Deviations, and Ranges for Dictation Test Error Scores (N's=26)

Recordings were made in Colorado. Tests were administered at 80 wpm.

Variable No.	Test Form	Order of Presentation	Method of Presentation	M	SD	Range
1	5 (Three	First equated groups)	Live	7.1	5.7	0-22
2	10	Second	Live	11.0	10.9	0-44
3	10	Second	Earphones	11.9	9.2	0-37
4	10	Second	Loudspeaker	16.0	15.7	0-71

TABLE III-2
(Colorado)

Correlation Coefficients between Dictation Test Error Scores for Test No. 5 Administered Live First and Test No. 10 Administered Second by Three Methods: Live, Earphones and Loudspeaker (N's=26)

Recordings were made in Colorado. Tests were administered at 80 wpm.

Variable Numbers	Test and Methods	r
1 and 2	Test 5 live and Test 10 live	.83**
1 and 3	Test 5 live and Test 10 earphones	.42*
1 and 4	Test 5 live and Test 10 loudspeaker	.76**

* Significant at the .05 level
**Significant at the .01 level

STUDY IV
(Utah)

Summary

Study IV was conducted by the Utah employment security agency in 1966 to compare reliability coefficients for live and recorded dictation test scores. Examinees' preferences for live and recorded methods and for male and female voice were also studied. The sample was composed of 216 stenography students and 26 local office applicants. All subjects took Form No. 4 first and Form No. 10 second at 80 wpm. One group of 119 subjects took the live presentation first and the recorded second. Another group of 123 subjects took the recorded method first and the live method second. For the recorded method, the male voice of an experienced radio announcer was put on tape by professional sound engineers in Utah. Tapes were used with loudspeakers. Demographic data and preference data were reported at the time of testing. The reliability coefficient between error scores on different test forms for 119 subjects taking the tests live was .75. For 123 subjects taking recorded tests the correlation was .84. The recorded and live methods were equally preferred. Most examinees preferred a male voice.

The Utah study contains valuable information on demographic variables, test conditions, and preference data and will be published separately as USTES Test Research Report No. 28-a.

Purpose

1. To compare reliability coefficients for live and recorded dictation scores.

2. To study examinees' preferences for live and recorded dictation methods.

3. To study examinees' preferences for male and female voice for dictation test administration.

Procedure

Sample. The total sample of 242 examinees included 216 shorthand students at Brigham Young University and 26 local office applicants. There were two male subjects in the sample. The average subject was 20 years old, had 13 1/2 years of education, had taken two courses in shorthand, and had approximately 6 months shorthand work experience. Ten subjects who did not attempt all of the

transcription were deleted from the sample. Three subjects were eliminated because demographic data was incomplete.

Test Administration. Forms No. 4 and No. 10 were administered at 80 wpm. to 242 subjects. One group of 119 subjects took both tests live; another group of 123 subjects took both tests recorded. The sequence of forms administered was the same for both groups--Form No. 4 first and Form No. 10 second. Recorded tests were administered by using loudspeakers and tape recordings of the male voice of an experienced radio announcer. The recordings were made by professional sound engineers in Utah.

Two male and two female examiners participated in the study.

Because test administration for students had to be limited to a 50 minute class period, no break was allowed between tests and standard transcription time was reduced from 20 minutes to 15 minutes for both tests.

At the time of testing, subjects completed an information sheet on background data. Subjects were asked to report their preference of method of test administration and sex of examiner. Categories for reporting preferences on method were "live", "recorded", and "either".

Categories for recording preferences on sex of examiner were "male", "female", and "either".

Variables.

Demographic data (N=242)

1. Age (years)
2. Education (years)
3. Training (no. of courses)
4. Experience (months)
5. Sex

Test data (N=242)

6. Test form (No. 4, No. 10)
7. Test method (N=119 live; N=123 recorded)
8. Sex of examiner
9. Number of examiners
10. Error scores

Preference data (N=241)

11. Method (live, recorded, either)
12. Sex of examiner (male, female, either)

Statistical Treatment. Quantitative analysis included computation of means, standard deviations, ranges, Pearson produce-moment correlations, standard errors of the differences between means and between reliability coefficients, analysis of variance, t ratios and F ratios.

Results and Discussion

Reliability coefficients between dictation test error scores on different test forms for live and recorded methods with accompanying means and standard deviations are shown in Table IV-1. The r's of .75 for live socres and .84 for recorded scores are not significantly different. All subjects in the sample had scores within the range of accuracy norms for employed workers. Therefore, both methods are reasonably reliable for persons who are proficient enough to be employed.

Data in Table IV-2 show that both methods of administration were equally preferred and that 41% of the total sample expressed no preference for either the live or the recorded method. Data in Table IV-3 show a preference for a male voice.

The Utah study contains many statistical analyses on reliability and preference data for subsamples. In general, the results of subsample analyses support the findings for the total sample. The original Utah report will be published as USTES Test Research Report No. 28-a.

Conclusions

The data reported for the Utah sample indicate that:

1. Live and recorded methods of test administration are equally reliable.

2. Live and recorded methods of test administration are equally preferred.

3. The male voice is preferred for dictation test administration.

TABLE IV-1
(Utah)

Means, Standard Deviations and Correlation Coefficients for Dictation Test Error Scores Obtained by Live and Recorded Methods*

Tape recordings were professionally produced in Utah using a trained male voice and administered by using loudspeakers.

Method	N	Error Score on First Test (No. 4)		Error Score on Second Test (No. 10)		r 12
		M	SD	M	SD	
Live	119	11.7	13.6	12.8	15.4	.75
Recorded	123	12.1	12.8	13.0	13.4	.84

+Table IV-1 is reprinted from a paper presented at a meeting of the American Psychological Association (Crambert, 1968).

*The sample included only subjects who attempted all of the transcription. Scores ranged from 0-75 errors. Tests were administered at 80 wpm.

TABLE IV-2
(Utah)

Preferences in Percents for Method of Dictation Test Administration (N=241)

	Live	Recorded	Either
Preferences in Percents	29	30	41

TABLE IV-3
(Utah)

Preferences in Percents for Administration of Dictation Tests by Male or Female Voice (N=241)

	Male	Female	Either
Preferences in Percents	70	5	25

STUDY V
(Alabama)

Summary

This study was conducted by the Alabama employment security agency in 1966 to compare the reliability of dictation test scores administered by live and recorded methods and to survey a sample of Southern examinees' opinions about recorded dictation produced in a non-Southern State. The sample was composed of students, local office applicants, and employed workers. Test Forms No. 10 and No. 11 were administered, always in that order, at 80 wpm to one group live and another group recorded. When data were analyzed for reliability study, subjects with more than 88 errors on the first test were deleted from the analyses. The recordings were produced in Colorado and used in Alabama with loud-speakers. Subjects who took dictation by the recorded method were asked to reply "yes" or "no" to three questions about recordings. Results of the survey show that 54% of the group expressed an overall favorable reaction to the recorded method and that over 80% of the Alabama subjects thought that sound production and pronunciation of words were satisfactory. Widely discrepant error score data for the same persons on test and retest make any interpretation of statistical findings on test-retest reliability for live and recorded methods of test administration questionable.

Purpose

(1) To compare the reliability of dictation test error scores when tests are administered by live and recorded methods

(2) To survey opinions on live and recorded methods in a Southern State for recordings produced in a non-Southern State

Procedure

Sample. The original sample of 209 subjects consisted of 139 stenography students, 31 local office applicants, 25 employed stenographers, and 14 MDTA stenography students. The total sample had an average age of 22.5 years and an average education of 13 years. There were 27 subjects with more than 88 errors on the first test who were deleted from the analyses on test reliability.

Test Administration. All subjects took Form No. 10 first and Form No. 11 second at 80 wpm. There were 93 subjects who took the tests live

and 89 subjects who took both tests recorded. Phonographs and loudspeakers were used to play recordings produced in Colorado. Subjects who took dictation by the recorded method were asked to answer three questions about the recorded dictation. Data were collected on age and education.

Variables.

Demographic (N=209)

1. Age (years)
2. Education (years)

Dictation Test Error Scores

3. Form No. 10, live (N=93)
4. Form No. 11, live (N=93)
5. Form No. 10, recorded (N=89)
6. Form No. 11, recorded (N=89)

Questions -- yes or no response (N=105)

7. Do you feel that the sound production was satisfactory?
8. Do you feel that the pronunciation of words was clear?
9. Is your overall reaction favorable to this type of test (recorded)?

Statistical Treatment. Means, standard deviations, and Pearson product-moment correlations were computed for error scores. Frequency data and percents were computed for opinion survey data.

Results and Discussion

The live and recorded groups were similar on age and education.

Reliability coefficients are .72 for both live and recorded scores and are shown with accompanying means and standard deviations in Table V-1. Live and recorded means are similar for first test administered and for second test administered. However, the differences between means for the first test and means for second test administered are extremely large. The distributions for the first test are positively skewed with high concentration of subjects having 0-9 errors. The distribution for the second test are platykurtic and asymmetric. For equivalent forms, the results are very atypical.

The percent figures for subjects answering "yes" to three questions are shown in Table V-2. Most subjects in the Alabama sample thought that reproduction of sound and pronunciation of words were satisfactory for the Colorado recording. An overall favorable reaction to the recorded method was marked "yes" by 54% of the group surveyed.

Conclusions

Equal reliability coefficients high enough to indicate dependable test-retest results for both live and recorded methods are reported in the study. However, the difference between test score means and distributions on presumably equivalent forms prohibits conclusive interpretation of findings.

Of 105 Alabama subjects who took recorded dictation produced in Colorado, 54% reported an overall favorable reaction to the recorded method, and over 80% thought that sound reproduction and pronunciation of words were satisfactory.

TABLE V-1
(Alabama)

Means, Standard Deviations and Correlation Coefficients for Dictation Test Error Scores Obtained by Live and Recorded Methods*

Recordings were made in Colorado and administered by using loudspeakers.

Method	N	Error Score on First Test (No. 4)		Error Score on Second Test (No. 11)		r_{12}
		M	SD	M	SD	
Live	93	26.8	24.6	62.9	39.4	.72
Recorded	89	27.6	23.4	60.9	39.8	.72

*Subjects with more than 88 errors on the first test were deleted from the analysis. Tests were administered at 80 wpm.

TABLE V-2
(Alabama)

Percent of Alabama Subjects in the Recorded Dictation Test Group Who Answered "Yes" to Three Questions (N=105)*

Recordings were made in Colorado and administered with loudspeakers.

Question	Percent "Yes"
1. Do you feel that the sound reproduction was satisfactory?	89
2. Do you feel that the pronunciation of words was clear?	82
3. Is your overall reaction favorable to this type test (recorded)?	54

*The sample included subjects with scores ranging from 0-178 errors.

STUDY VI
(Minnesota)

Summary

This study was conducted by the Minnesota employment security agency in 1966 to study differences in score level for live and recorded methods of dictation test administration and to study examinees' preferences for method of administration. The sample of 100 female subjects included local office applicants, students and employed workers. Subjects making more than 50 errors on the initial test were eliminated from the sample. Forms No. 10 and No. 4 were administered in that order at 80 wpm. to two groups each with 50 subjects. Group I took recorded dictation first and live dictation second. Group II took live dictation first and recorded dictation second. Tape recordings were made in Minnesota and used with loudspeakers. Each examinee completed a background questionnaire which included a multiple-choice question on preference for live or recorded administration. Five alternative responses allowed subjects to express degree of preference. There were no significant differences in mean error score level for the total sample taking the tests by live and recorded methods. Over two-thirds of the sample preferred live administration, but only 30% of the subjects strongly preferred live administration.

Purpose

(1) To study differences in score level for live and recorded methods of dictation test administration.

(2) To assess personal preferences of examinees for each method.

Procedure

Sample. The 100 subjects were female. Forty-five were local office applicants; four were MDTA stenographic students; and 51 were employed workers in civil service and private industry. Any subject making more than 50 errors on the first test was eliminated from the sample.

Test Administration. Forms No. 4 and No. 10 were administered at 80 wpm. Form No. 10 was always administered first. The sample was divided into two groups with 50 subjects in each one. Group I took recorded dictation first and live dictation second. Group II took live dictation first and recorded dictation second. There were two tapes used.

A different dictator was recorded on each tape. Tapes were used with loudspeakers. Several test administrators participated in the study because there are a variety of administrators in the operational situation. There was one exception to standard testing procedure. No practice exercise was administered before the second test administration. At the time of testing, each examinee completed a background questionnaire which included one multiple-choice question on preference for live or recorded dictation. The five alternative responses allowed subjects to express degree of preference.

Variables.

Demographic data (N=100)

1. Age (years)
2. Education (years)
3. Experience (months)

Dictation Test Error Scores (N=50)

4. Group I, Form 10, recorded first
5. Group I, Form 4, live second
6. Group II, Form 10, live first
7. Group II, Form 4, recorded second

Preference data (N=100) -- alternative responses on test administration method

8. Does not matter
9. Strongly prefer recorded
10. Mildly prefer recorded
11. Strongly prefer live
12. Mildly prefer live

Statistical Treatment. Computations included means, standard deviations, and Pearson product-moment coefficients of correlation. Differences between independent means were tested by applying the z ratio for uncorrelated means, and differences between test-retest means were treated by applying the z ratio for correlated means (Underwood, et. al., 1954). In view of the positive skew in all error score distributions, error score means were also compared by applying two nonparametric measures: the median test by chi-square and the Wilcoxin test of signed ranks of differences between matched pairs (Peatman, 1963). A third nonparametric technique, chi-square analysis (Garrett, 1958), was applied

to frequency data on preferences. The interaction effect between preferences and being above or below the median on age, education, experience, or combined error score was also treated by chi-square analysis (McNemar, 1955).

Results and Discussion

Demographic data and correlations between demographic variables and combined live and recorded error scores for 100 cases are shown in Table VI-1. The correlation of .63 between live and recorded error scores was considered high enough to combine error scores. The average subject was 26 years old with 12 years of education and 39 months experience. The negative correlation between experience and errors was significantly different from zero at the .05 level.

Means on age, education, experience, and combined error score for the two experimental groups were computed and tested for significant differences. No significant differences were found between Groups I and II on these four variables.

Means for total live scores and total recorded scores are shown in Table VI-2. The distributions of live and recorded scores were not significantly different on any of the tests applied.

All of the group means in Table VI-2 were compared by several statistical methods. Significant differences were found between first and second tests administered. However, the relative size of the means appears to relate to position in administration sequence and not to method of administration. Means for the two second test administrations were significantly different according to the median test by chi-square. A possible explanation for this is that Group II did not receive practice with the recorder prior to the actual test -- a condition which might account for increased errors on the recorded test for Group II.

Preference data are shown in Table VI-3. The preferences for the live method for each group and for the total sample were significant at the .001 level. Interaction effects between preferences and four variables (age, education, experience, and combined error scores) were also studied. No significant differences on preference data were found between high and low groups on the four variables studied.

TABLE VI-1
(Minnesota)

Means, Standard Deviations and Ranges for Demographic Variables and Correlation Coefficients between Demographic Variables and Combined Live and Recorded Dictation Test Error Scores for 100 Subjects

Demographic Variable	M	SD	Range	r (demographic data and combined error score +)
Age (years)	26.0	10.8	18-64	-.18
Education (years)	12.3	.8	12-16	.09
Experience (months)	39.2	66.5	0-360	-.20*

+The correlation between live and recorded error scores for 100 cases was .63.

*Significant at the .05 level.

TABLE VI-2
(Minnesota)

Means and Standard Deviations for Dictation Test Error Scores Obtained on First and Second Administrations by Live and Recorded Methods and for Live and Recorded Combined Samples*

Tape recordings were made in Minnesota and used with loudspeakers.

	Error Score, First Test Form No. 10 (N=50)			Error Score, Second Test Form No. 4 (N=50)			All Live (N=100)		All Recorded (N=100)	
Group	Method	M	SD	Method	M	SD	M	SD	M	SD
I	Recorded	17.6	20.2	Live	10.6	12.6	13.7	14.7	16.2	17.1
II	Live	16.9	15.9	Recorded	14.9	13.2				

*Subjects making more than 50 errors on the first test were deleted from the sample. Tests were administered at 80 wpm.

TABLE VI - 3
(Minnesota)

Frequency Data and Percents for Preferences on Live and Recorded Test Administration for Groups I and II*

Tape recordings were made in Minnesota and used with loudspeakers.

Preference	Group I (recorded-live) N=50		Group II (live-recorded) N=50	
	N	%	N	%
Does not matter	13	26	4	8
Strongly prefer recorded	3	6	2	4
Mildly prefer recorded	1	2	8	16
Strongly prefer live	20	40	15	30
Mildly prefer live	13	26	21	42
Total	50	100	50	100

*Subjects were limited to those with 50 errors or less on the first dictation test.

Conclusions

(1) Differences in mean score level do not relate to method of test administration.

(2) Over two-thirds of the examinees had a mild or a strong preference for the live method.

STUDY VII
(Mississippi)

Summary

This study was conducted by the Mississippi employment security agency in 1966 to determine whether score level for live and recorded administrations is equivalent and to survey preferences of Southern examinees for live and recorded dictation. The sample consisted of 123 stenographic students. Data were recorded for age, education, experience, training, and previous experience with recorded dictation and with USES Dictation Tests. Forms No. 10 and No. 4 were administered in that order at 80 wpm to two groups. Group I took live administration first and recorded second. Group II took recorded administration first and live second. Tapes had been recorded in Minnesota and were used with loudspeakers. Test score data were treated for the total range of scores (0-196 errors), for the scorable range of the accuracy norms (0-88 errors), and for norm grades A, B, and C (0-14 errors). After testing, examinees expressed preference for method of administration by checking one of five alternative statements which included statements on degree of preference. Results of statistical treatment showed that good stenographers performed better on tests administered by the live method. Method of administration made no difference in the scores of poor stenographers. Almost three-fourths of the total sample preferred live administration. However, only 29% had a decided preference for live administration. The Mississippi figures on preferences are very similar to those shown for Minnesota.

Purpose

(1) To determine whether score level is equivalent for live and recorded administration

(2) To determine by a preference survey whether Southern subjects had any negative reaction to recorded dictation produced in a non-Southern State

Procedure

Sample. The sample consisted of 123 female students in advanced stenography classes at seven State junior colleges in Mississippi. Most individuals in the sample were also local office applicants. Examinees were informed that test results would be used for research but were not informed that recorded tests were for research purposes only until after the testing session. No member of the sample had taken either of the

specific test forms used in this study although 34 subjects had taken the USES Dictation Test previously.

Test Administration. Test Forms No. 10 and No. 4 were administered at 80 wpm to two groups. Form No. 10 was always administered first. Group I took the live method first and recorded method second. Group II took the recorded method first and the live method second. A one-minute recorded practice exercise preceded each recorded test administration; a one-minute live practice exercise preceded each live administration. The tapes had been recorded by the Minnesota State Employment Service and were used in Mississippi with loudspeakers. One recording was on acetate tape and the other was on tenzar tape. Timing was checked both before and after the study. On each check both test forms exceeded the standard three-minute time by six seconds. No technical difficulties were encountered.

At the time of testing, information on age, education, experience, training, and previous experience with the USES Dictation Test and with recorded material was recorded. After testing, examinees checked one of five alternative statements on preference and degree of preference for live or recorded dictation.

Test data were treated for the total range of scores, the scorable range of the accuracy norms, and the A, B, and C norm grades combined.

Variables.

Demographic (N=123)

1. Age (years)
2. Education (years)
3. Experience (months)
4. Training (years)
5. Previous experience taking dictation from recorded material (yes-no)
6. Previous experience taking a USES Dictation Test (yes-no)

Error Scores

7. Total sample -- 0-196 errors (N=123)
8. Group I -- 0-196 errors (N=62)
9. Group II -- 0-105 (N=61)
10. Total sample -- 0-88 errors (N=111)
11. Total sample -- 0-14 errors (N=64)
12. Total sample -- live score range 0-14 errors (N=59)

Preference variables; five alternative responses:

13. "I wouldn't care which was used - both are about the same."
14. "I would definitely prefer the recorded dictation."
15. "I would prefer the recorded dictation, but it wouldn't make much difference."
16. "I would definitely prefer the 'live' dictation."
17. "I would prefer the 'live' dictation, but it wouldn't make much difference."

Statistical Treatment. Computations included means, standard deviations, t ratios for the standard error of the difference between correlated means, Pearson product-moment correlation coefficients and point biserial coefficients.

Results and Discussion

Descriptive data for the sample and for two groups are shown in Table VII-1. The average subject was approximately 19 years old with 13 years of education, one month stenography experience, and two years of stenography training. The two groups had similar means on four demographic variables.

Frequency data and percents for previous experience with recorded dictation and with the USES Dictation Test are shown in Table VII-2. Point biserial coefficients of correlation between recorded dictation test scores and these two variables were not significantly different from zero. The same computations for the subsample with 0-14 errors were not statistically significant.

Test data are shown in Table VII-3. The difference between live and recorded means for the full range of scores regardless of order of method presentation was not significant at the .05 level, and correlations between live and recorded scores were high. Means were not significantly different for the scorable range of errors (0-88), and the correlation between live and recorded scores for this sub-group was .83. However, for 58% of the sample who would have had grades A-C (0-14 errors) on either test and for 53% of the sample who would have had grades A-C on the live test, the difference between live and recorded means was significant at the .01 level. Pearson r's between live and recorded error scores for these two subsamples were .33 and .61.

TABLE VII-1
(Mississippi)

Means and Standard Deviations for Demographic Variables for the Total Sample and Two Experimental Groups

		Age (years)		Education (years)		Experience (months)		Training (years)	
Group	N	M	S.D.	M	S.D.	M	S.D.	M	S.D.
Total	123	18.7	1.0	13.2	.4	1.3	3.2	2.4	.6
I	62	18.8	1.2	13.3	.5	1.5	2.6	2.3	.7
II	61	18.6	.8	13.1	.3	1.1	3.7	2.4	.6

TABLE VII-2
(Mississippi)

Number and Percent of Subjects Having Previous Experience Taking Dictation from Recordings and Taking USES Dictation Test (N=123)

Previous Experience	Number	Percent
Taking dictation from recordings	96	78.0
Taking USES Dictation Test	34	27.6

TABLE VII-3
(Mississippi)

Means, Standard Deviations, t Ratios for the Standard Error of the Difference between Means, and Correlation Coefficients between Live and Recorded Dication Test Error Scores for the Total Sample, Two Experimental Groups, and Three Subsamples Defined by Error Score Range*

Recordings were made in Minnesota and used with loudspeakers.

Group (Range of errors)	N	Live Error Scores M	Live Error Scores SD	Recorded Error Scores M	Recorded Error Scores SD	t Ratio	r live & recorded
Total (0-196)	123	28.8	33.3	30.9	31.5	1.43	.89
Group I (0-196)	62	35.8	38.8	37.3	38.5	.70	.91
Group II (0-105)	61	21.7	23.1	24.4	22.3	1.55	.82
0-88 Errors, Either Test	111	21.1	20.7	23.1	19.6	1.71	.83
0-14 Errors, Either Test	64	6.3	5.5	9.6	8.4	3.07**	.33
0-14 Errors, Live Test	59	5.2	3.4	9.6	8.7	4.73**	.61

* Form No. 10 was always administered first; Form No. 4 was administered second. Both tests were given at 80 wpm.

**Significant at the .01 level.

TABLE VII-4
(Mississippi)

Number and Percent of Responses for Each Preference Alternative for 123 Dictation Test Examinees in Mississippi

Examinees had heard recordings made in Minnesota and used with loudspeakers.

Preference	Number	Percent
None	19	15.4
Recorded	7	5.7
Recorded with little difference	6	4.9
Live	36	29.3
Live with little difference	55	44.7
Total	123	100.0

Table VII-4 shows survey results. Although 74% of the examinees preferred live dictation, only 29% had a decided preference for the live method. No preference was expressed by 15% of the sample. These figures are very similar to those reported in Study VI for the Minnesota sample. In the Minnesota study, 35% of the total sample had a strong preference for live dictation and 34% had a mild preference for the live method. Seventeen percent of the Minnesota subjects had no preference.

Test administrators collecting data for Study VII in Mississippi had no adverse reactions to the use of recordings. They reported that examinees in this study did not show any differences between their reactions to recordings and to live tests.

Conclusions

The data reported by the Mississippi agency indicate that:

(1) Good stenographers take dictation more accurately on tests administered by the live method;

(2) Method of administration makes no difference in average score level for poor stenographers;

(3) The majority of examinees preferred live dictation.

(4) Subjects in Mississippi and in Minnesota who heard the same recordings produced in Minnesota had very similar preference responses.

Summary of Findings for the Seven Studies

The bulk of the data from these studies supports the hypothesis that the recorded administration of the 1953 USES Dication Tests produces essentially the same results, in terms of mean scores and array of scores, as the live administration of the same tests. In general, the correlations between the recorded and live administrations were high enough to consider the two methods of administration to be equivalent in terms of ranking persons on their ability to take dictation.

Phonograph recordings and tape recordings produced similar results although no formal comparison was made. Since there were procedural or sampling difficulties in the studies in which earphones were used, the question of the comparative utility of loudspeakers and earphones was not answered. However, several examinees complained that the earphones were a nuisance.

The question of examinee preference was not clearly answered. Two States reported no preference and two States reported some preference or strong preference for live administration.

The usefulness of recordings made in one geographic region for dictation in another region with different speech patterns was supported. The data was limited, however, to the use of recordings made in States not located in the Southeastern United States to southern States.

Recommendations

Although the five questions on live and recorded methods listed in the Introduction were not all unambiguously answered by the seven separate studies, no observations were made that would prohibit using the recorded method of test administration for dictation tests. In 1969, the Utah State Employment Service had the new true-false, separate answer sheet USTES Dictation Test (Forms A-F) recorded at 60, 80, and 100 wpm. A trained male voice was selected and recordings were professionally produced. The new recordings will replace all of the experimental ones and will be available to all State employment security agencies.

Review of data in the seven comparative studies, however, raises certain important questions. For example, what are the practice and fatigue effect expected in the dictation test-retest situation? Subjects performed better on the second test administered in the New York and Minnesota studies. Poorer performance was shown on the second test administered in the Colorado, Utah, Alabama, and Mississippi studies. An important question about dictation test reliability is raised by the Mississippi study. Is dictation test reliability the same at high and low score levels? The difference in findings for good and poor stenographers on live and recorded methods of test administration may reflect a difference in test reliability at different score levels. The interaction of demographic variables with the practice and fatigue variables and at different levels of test scores has not been investigated thoroughly.

Therefore, it is recommended that reliability studies be done separately for live and recorded dictation methods using the new USTES Dictation Tests followed by an equivalence study for live and recorded methods. The recommended studies would have the advantage of results based on tests and recordings in present operational use.

Review of the seven comparative studies on live and recorded dictation test methods also indicates a number of procedural suggestions for future research on the USTES Dictation Tests, as follows:

1. Data on age, education, and experience should be collected for all subjects.

2. Scores on the new USTES Spelling Test should be available for all subjects.

3. Experimental groups should be similar on demographic variables and USTES Spelling Test scores. They should not be matched on dictation test scores.

4. There should be arbitrary standards set for research purposes to insure that subjects included in the sample are able to attempt dictation at a given speed. No subject should be included in the sample who does not attempt all of the transcription on the first dictation test administered.

5. Data should be treated separately for the full range of scores and for high and low scoring subsamples.

6. Data should be available on equivalence of test forms used in research studies.

References

Crambert, Albert C. Differences in examinee performance on "live" and recorded dictation tests. Paper presented at the meeting of the American Psychological Association, San Francisco, 1968.

Edwards, Allen L. Statistical analysis. New York: Rinehart and Co., 1951.

Garrett, Henry E. Statistics in psychology and education. New York: Longmans, Green and Co., 1958.

McNemar, Q. Psychological statistics. New York: John Wiley and Sons, Inc., 1955.

Peatman, J. G. Introduction to applied statistics. New York: Harper and Row, 1963.

Siegil, Sidney. Nonparametric statistics for the behavioral sciences. New York: McGraw-Hill Book Co., Inc., 1956.

Underwood, B. J.; Duncan, C. P.; Taylor, J. A.; and Cotton, J. W. Elementary statistics. New York: Appleton-Century-Crofts, Inc., 1954.

U. S. Department of Labor, Bureau of Employment Security. Guide to the use of Typing, Dictation and Spelling Tests. Washington, D. C.: U.S. Government Printing Office, December, 1953.

U. S. Department of Labor, Manpower Administration. Manual for USES Clerical Skills Tests. Washington, D. C.: U. S. Government Printing Office, March, 1968.

Walker, Helen M. and Lev, Joseph. Statistical inference. New York: Henry Holt and Co., Inc., 1953.